

# Treinamento e síntese de FCNs com campo receptivo variável

Uma aplicação em segmentação de imagens

Pedro Pimentel

Departamento de Ciência da Computação — Universidade de São Paulo

pedro.pimentel@usp.br



IME-USP

## Introdução

A segmentação de imagens consiste na tarefa computacional de atribuir um rótulo para todo pixel pertencente a uma imagem. Este processo geralmente tem como objetivo diminuir a quantidade de informação mantendo o conteúdo principal da imagem, de forma a facilitar a interpretação ou tornar a informação intelectualmente mais significativa. Como exemplos podemos citar as áreas de interpretação de imagens médicas [2], e detecção de faces [1].

## Motivação

Tradicionalmente modelos computacionais utilizam Redes Neurais Convolucionais (CNNs) para realizar essa tarefa, com arquiteturas no estilo *bottle-neck*.

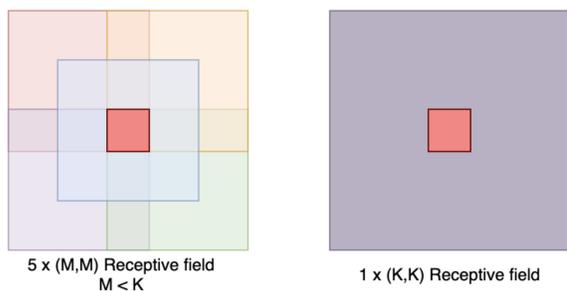


Figura 1: Escolha de campo receptivo. À direita um campo receptivo único de tamanho  $(K \times K)$  e objetivo de tamanho  $(O \times O)$ , à esquerda 5 campos receptivos de tamanho  $M < K$  e objetivo de tamanho  $(O \times O)$

Na maioria dos casos estes modelos consideram um campo receptivo fixo, com kernel no centro das imagens selecionadas. A variação entre posições de kernel e combinações de FCNs diferentes é um campo pouco explorado e que pode trazer bons desempenhos em determinados problemas.

## Objetivos

1. Explorar o impacto da escolha do campo receptivo na performance de modelos de segmentação de imagens.
2. Utilizar a arquitetura de Redes Neurais Totalmente Convolucionais (FCNs) para segmentação imagens.
3. Propor uma Rede Neural de síntese para as imagens produzidas pelas redes FCN.

## Metodologia

Utilizamos duas redes neurais para realizar a tarefa, separadas em **segmentação** e **síntese**. Na etapa de segmentação utilizamos modelos na arquitetura FCN. Cada modelo captura um campo receptivo diferente, compartilhando um mesmo *kernel* como conjunto objetivo.

Na etapa de síntese utilizamos uma rede neural densa para aprender a combinar os resultados dos diferentes modelos FCN de forma ótima.



Figura 2: Ilustração da arquitetura final desejada

## Redes Neurais Totalmente Convolucionais (FCNs)

Cada camada de dados em uma FCN é um vetor tridimensional de tamanho  $h \times w \times d$  onde  $h$  e  $w$  são dimensões espaciais e  $d$  é uma feature ou dimensão de canal. A primeira camada é a imagem, com tamanho de pixel  $h \times w$  e  $d$

canais de cor. Estas camadas são operadores locais, e dependem unicamente de suas coordenadas espaciais.

Definindo  $x_{ij}$  como o vetor da posição  $(i, j)$  de uma camada particular e  $y_{ij}$  como o vetor da camada subsequente, estes operadores locais computam  $y_{ij}$  como:

$$y_{ij} = f_{ks}(\{x_{si+\delta i, sj+\delta j}\}_{0 \leq \delta i, \delta j \leq k})$$

onde  $k$  é o tamanho do kernel e  $s$  é o *stride*.

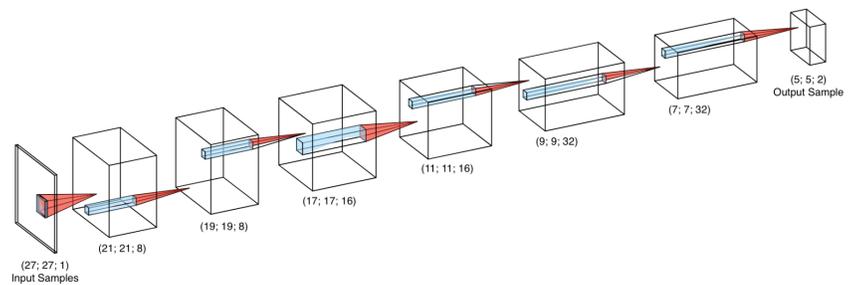


Figura 3: Arquitetura FCN utilizada.

## Rede Neural de Síntese

Para a etapa de síntese utilizamos uma rede neural densa como mostra a figura a seguir:

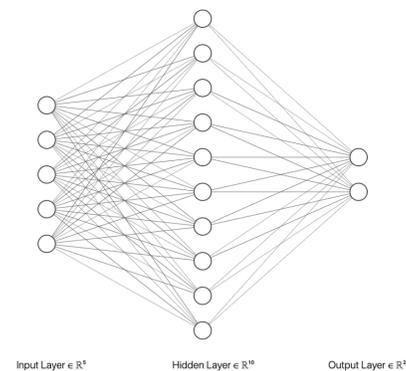


Figura 4: Arquitetura de Rede Neural Densa

A entrada desta rede consiste de um vetor  $x_{ij} \in R^5$  contendo as saídas das redes FCN para o pixel na posição  $(i, j)$ . Seu treinamento mapeia  $f : R^5 \rightarrow R^2$ . Uma vez treinado, o modelo foi aplicado em batch para as diferentes imagens obtidas com as FCNs.

## Resultados



Figura 5: Resultado do modelo: Na ordem: entrada, resultado desejado e resultado obtido

## Referências

[1] Nidhal El abbadi. Detection and segmentation of human face. *International Journal of Advanced Research in Computer*

and Communication Engineering, 4:90–94, 02 2015.

[2] SP. Chokkalingam N. Deepa. Deep convolutional neural networks for medical image analysis. *IJEAT*, 8(52), February 2019.